WHAT IS CLAIMED IS:

1. A method of identifying a set of informative genes or markers for a condition comprising a plurality of phenotypic or genotypic characteristics, comprising:

(a) classifying each of a plurality of samples or individuals on the basis of one or more phenotypic or genotypic characteristics of said condition into a plurality of first classes; and

(b) identifying within each of said first classes a first set of genes or markers informative for said condition

wherein said first set of genes or markers within each of said first classes is unique to said class relative to other first classes.

2. The method of claim 1, which further comprises additionally classifying into a plurality of second classes said samples or individuals in at least one of said first classes on the basis of a phenotypic or genotypic characteristic different that that used in said classifying step (a); and identifying within at least one of said second classes a second set of informative genes or markers, wherein said second set of informative genes or markers within each of said second classes is unique to said second class relative to other first and second classes.

3. A method of identifying a set of informative genes or markers for a condition comprising a plurality of phenotypic or genotypic characteristics, comprising:

(a) classifying each of a plurality of samples or individuals on the basis of one or more phenotypic or genotypic characteristics into a plurality of first classes;

(b) classifying at least one of said first classes into a plurality of second classes on the basis of phenotypic or genotypic characteristic different than that used in said classifying step (a); and

(c) identifying within at least one of said first classes or said second classes a set of genes or markers informative for said condition,

wherein said second set of genes or markers is unique to said class relative to other first and second classes.

4. A method of identifying a set of informative genes or markers for a condition comprising a plurality of phenotypic or genotypic characteristics, comprising:

(a) selecting a first characteristic from said plurality of phenotypic or genotypic characteristics;

(b) identifying at least two first condition classes differentiable by said first characteristic;

(c) selecting a plurality of individuals classifiable into at least one of said first condition classes; and

(d) identifying in samples derived from each of said plurality of individuals a set of genes or markers informative for said condition within said at least one of said first condition classes.

5. A method of classifying an individual with a condition as having a good prognosis or a poor prognosis, comprising:

(a) classifying said individual into one of a plurality of patient classes, said patient classes being differentiated by one or more phenotypic, genotypic or clinical characteristics of said condition;

(b) determining the level of expression of a plurality of genes or their encoded proteins in a cell sample taken from the individual relative to a control, said plurality of genes or their encoded proteins comprising genes or their encoded proteins informative for prognosis of the patient class into which said individual is classified; and

(c) classifying said individual as having a good prognosis or a poor prognosis on the basis of said level of expression.

6. The method of claim 5, wherein said condition is cancer, said good prognosis is the non-occurrence of metastases within five years of initial diagnosis, and said poor prognosis is the occurrence of metastases within five years of initial diagnosis.

7. The method of claim 5, wherein said control is the average level of expression of each of said plurality of genes or their encoded proteins across a plurality of samples derived from individuals identified as having a poor prognosis.

8. The method of claim 7, in which said classifying step (c) is carried out by a method comprising comparing the level of expression of each of said plurality of genes or their encoded proteins to said average level of expression of each corresponding gene or its encoded protein in said control, and classifying said individual as having a poor prognosis if said level of expression correlates with said average level of expression of each of said genes or their encoded proteins in said control more strongly than would be expected by chance.

9. The method of claim 5, wherein said control is the average level of expression of each of said plurality of genes or their encoded proteins across a plurality of samples derived from individuals identified as having a good prognosis.

10. The method of claim 9, in which said classifying in step (c) is carried out by a method comprising comparing the level expression of each of said plurality of genes or their encoded proteins to said average level of expression of each corresponding gene or its encoded protein in said control, and classifying said individual as having a good prognosis if said level of expression correlates with said average level of expression of each of said genes or their encoded proteins in said control more strongly than would be expected by chance.

11. The method of claim 5, wherein said plurality of patient classes comprises ER^-, *BRCAI* individuals; ER^-, sporadic individuals; ER+, ER/AGE high individuals; ER+, ER/AGE low, LN+ individuals; and ER+, ER/AGE low, LN^- individuals.

12. A method of classifying a breast cancer patient as having a good prognosis or a poor prognosis comprising:

(a) classifying said breast cancer patient as ER^-, *BRCAI*; ER^-, sporadic; ER+, ER/AGE high; ER+, ER/AGE low, LN+; or ER+, ER/AGE low, LN^-;

(b) determining the level of expression of a first plurality of genes in a cell sample taken from said breast cancer patient relative to a control, said first plurality of genes comprising two of the genes corresponding to the markers in Table 1 if said breast cancer patient is classified as ER^-, *BRCAI*; in Table 2 if said breast cancer

patient is classified as ER^- sporadic; in Table 3 if said breast cancer patient is classified as ER+, ER/AGE high; in Table 4 if said breast cancer patient is classified as ER+. ER/AGE low, LN+; or in Table 5 if said breast cancer patient is classified as ER+, ER/AGE low, LN^-; and

(c) classifying said breast cancer patient as having a good prognosis or a poor prognosis on the basis of the level of expression of said first plurality of genes,

wherein said breast cancer patient is "ER/AGE high" if the ratio of the $\log_{10}$(ratio) of ER gene expression to age exceeds a predetermined value, and "ER/AGE low" if the ratio of the $\log_{10}$(ratio) of ER gene expression to age does not exceed said predetermined value.

13. The method of claim 12, wherein said control is the average level of expression of each of said plurality of genes in a plurality of samples derived from ER^-, *BRCA1* individuals, if said breast cancer patient is ER^-, *BRCA1*; the average level of expression of each of said plurality of genes in a plurality of samples derived from ER^-, sporadic individuals if said breast cancer patient is ER^-, sporadic; the average level of expression of each of said plurality of genes in a plurality of samples derived from ER+, ER/AGE high individuals, if said breast cancer patient is ER+, ER/AGE high; the average level of expression of each of said plurality of genes in a plurality of samples derived from ER+, ER/AGE low, LN+ individuals where said breast cancer patient is ER+, ER/AGE low, LN+; or the average level of expression of each of said plurality of genes in a plurality of samples derived from ER+, ER/AGE low, LN^- individuals where said breast cancer patient is ER+, ER/AGE low, LN^-.

14. The method of claim 13, wherein each of said individuals has a poor prognosis.

15. The method of claim 13, wherein each of said individuals has a good prognosis.

16. The method of claim 14, wherein said classifying step (c) is carried out by a method comprising comparing the level of expression of each of said plurality of genes or their encoded proteins in a sample from said breast cancer patient to said control, and classifying said breast cancer patient as having a poor prognosis if said level of expression

correlates with said average level of expression of the corresponding genes or their encoded proteins in said control more strongly than would be expected by chance.

17. The method of claim 12, wherein said predetermined value of ER is calculated as ER = 0.1(AGE – 42.5), wherein AGE is the age of said individual.

18. The method of claim 12, wherein said individual is ER^-, *BRCA1*, and said plurality of genes comprises two of the genes for which markers are listed in Table 1.

19. The method of claim 12, wherein said individual is ER^-, *BRCA1*, and said plurality of genes comprises all of the genes for which markers are listed in Table 1.

20. The method of claim 12, wherein said individual is ER^-, sporadic, and said plurality of genes comprises two of the genes for which markers are listed in Table 2.

21. The method of claim 12, wherein said individual is ER^-, sporadic, and said plurality of genes comprises all of the genes for which markers are listed in Table 2.

22. The method of claim 12, wherein said individual is ER+, ER/AGE high, and said plurality of genes comprises two of the genes for which markers are listed in Table 3.

23. The method of claim 12, wherein said individual is ER+, ER/AGE high, and said plurality of genes comprises all of the genes for which markers are listed in Table 3.

24. The method of claim 12, wherein said individual is ER+, ER/AGE low, LN+, and said plurality of genes comprises two of the genes for which markers are listed in Table 4.

25. The method of claim 12, wherein said individual is ER+, ER/AGE low, LN+, and said plurality of genes comprises all of the genes for which markers are listed in Table 4.

26. The method of claim 12, wherein said individual is ER+, ER/AGE low, LN^-, and said plurality of genes comprises two of the genes for which markers are listed in Table 4.

27. The method of claim 12, wherein said individual is ER+, ER/AGE low, LN^-, and said plurality of genes comprises all of the genes for which markers are listed in Table 4.

28. The method of claim 12, further comprising determining in said cell sample the level of expression, relative to a control, of a second plurality of genes for which markers are not found in Tables 1-5, wherein said second plurality of genes is informative for prognosis.

29. A method for assigning an individual to one of a plurality of categories in a clinical trial, comprising:

(a) classifying said individual as ER^-, *BRCA1*, ER^-, sporadic; ER+, ER/AGE high; ER+, ER/AGE low, LN+; or ER+, ER/AGE low, LN^-;

(b) determining for said individual the level of expression of at least two genes for which markers are listed in Table 1 if said individual is classified as ER^-, *BRCA1*; Table 2 if said individual is classified as ER^-, sporadic; Table 3 if said individual is classified as ER+, ER/AGE high; Table 4 if said individual is classified as ER+, ER/AGE low, LN+; or Table 5 if said individual is classified as ER+, ER/AGE low, LN^-;

(c) determining whether said individual has a pattern of expression of said at least two genes that correlates with a good prognosis or a poor prognosis; and

(d) assigning said individual to one category in a clinical trial if said individual has a good prognosis, and assigning said individual to a second category in said clinical trial if said individual has a poor prognosis.

30. The method of claim 29, wherein said individual is additionally assigned to a category in said clinical trial on the basis of the classification of said individual as determined in step (a).

31. The method of claim 29, wherein said individual is additionally assigned to a category in said clinical trial on the basis of any other clinical, phenotypic or genotypic characteristic of breast cancer.

32. The method of claim 29, further comprising determining in said cell sample the level of expression, relative to a control, of a second plurality of genes for which markers are not found in Tables 1-5, wherein said second plurality of genes is informative for prognosis of breast cancer, and determining from the expression of said second plurality of

genes, in addition to said first plurality of genes, whether said individual has a good prognosis or a poor prognosis.

33. A method of identifying a set of genes informative for a condition, said condition having a plurality of phenotypic or genotypic characteristics such that samples may be categorized by at least one of said phenotypic or genotypic characteristics into at least one characteristic class, said method comprising:

(a) selecting a plurality of samples from individuals having said condition;

(b) identifying a first set of genes informative for said characteristic class using said plurality of samples;

(c) predicting the characteristic class of each of said plurality of samples;

(d) discarding samples for which said characteristic class is incorrectly predicted;

(e) repeating steps (c) and (d) at least once; and

(f) identifying a second set of genes informative for said characteristic class using samples in said plurality of samples remaining after step (e).

34. The method of claim 6, wherein said cancer is breast cancer.

35. A method for assigning an individual to one of a plurality of categories in a clinical trial, comprising:

(a) classifying the individual into one of a plurality of condition categories differentiated by at least one genotypic or phenotypic characteristic of the condition;

(b) determining the level of expression, in a sample derived from said individual, of a plurality of genes informative for said condition category;

(c) determining whether said level of expression of said plurality of genes indicates that the individual has a good prognosis or a poor prognosis; and

(d) assigning the individual to a category in a clinical trial on the basis of prognosis.

36. A method for identifying one or more sets of informative genes or markers for a condition in an organism, comprising:

(a) subdividing a plurality of individuals or samples derived therefrom of said organism subject to said condition into a plurality of classes based on one or more clinical, phenotypic or genotypic characteristics of said organism, wherein each said class consists of a plurality of individuals or samples derived therefrom of said organism each having said one or more clinical, phenotypic or genotypic characteristics specific for said class; and

(b) attempting to identify for each of one or more of said plurality of classes a set of genes or markers informative for said condition in individuals in said class,

wherein, if a set of genes or markers informative for said condition in individuals in said class is obtained for any of said one or more of said plurality of classes, said set of genes or markers is taken as a set of informative genes or markers for said condition in said organism.

37. The method of claim 36, further comprising, for each of one or more of said classes in which a set of genes or markers informative for said condition in individuals in said class cannot be obtained, repeating said steps (a) and (b) on said plurality of individuals or samples derived therefrom in said class such that said plurality of individuals or samples derived therefrom in said class is subdivided into a plurality of additional classes based on one or more clinical, phenotypic or genotypic characteristics of said organism which are different from those used for defining said class, wherein, for each of said plurality of additional classes, if a set of genes or markers informative for said condition in individuals in said class is obtained, said set of genes or markers is taken as a set of informative genes or markers for said condition in said organism.

38. A method for identifying one or more sets of informative genes or markers for a condition in an organism, comprising:

(a) subdividing a plurality of individuals or samples derived therefrom of said organism subject to said condition into a plurality of classes based on one or more clinical, phenotypic or genotypic characteristics of said organism, wherein each said class consists of a plurality of individuals or samples derived therefrom of said organism each having said one or more clinical, phenotypic or genotypic characteristics specific for said class;

(b) attempting to identify for each of one or more of said plurality of classes a set of genes or markers informative for said condition in individuals in said class, wherein if a set of genes or markers informative for said condition in individuals in said class is identified for any of said one or more of said classes, said set of genes or markers is taken as a set of informative genes or markers for a condition in said organism; and

(c) for each of one or more of said classes in which a set of genes or markers informative for said condition in individuals in said class cannot be obtained, repeating said steps (a) and (b) on said plurality of individuals or samples derived therefrom in said class such that said plurality of samples or individuals in said class is subdivided into a plurality of additional classes based on one or more clinical, phenotypic or genotypic characteristics of said organism which are different from those used those used for defining said class, wherein, for each of one or more of said plurality of additional classes, if a set of genes or markers informative for said condition in individuals in said class is obtained, said set of genes or markers is taken as a set of informative genes or markers for a condition in said organism.

39. The method of claim 38, wherein said condition is a type of cancer, and wherein each of said sets of genes or markers is informative of prognosis of individuals in a corresponding class.

40. The method of claim 39, wherein said condition is breast cancer, and wherein said one or more clinical, phenotypic or genotypic characteristics comprises age, ER level, ER/AGE, BRAC1 status, and lymph node status.

41. The method of claim 39, further comprising generating a template profile comprising measurements of levels of genes or markers of said set for said class representative of levels of the genes or markers in a plurality of patients having a chosen prognosis level.

42. A method for predicting a breast cancer patient as having a good prognosis or a poor prognosis, comprising:

(a) classifying said breast cancer patient into one of the following classes: (a1) ER^-, *BRCAI*; (a2) ER^-, sporadic; (a3) ER+, ER/AGE high; (a4) ER+, ER/AGE low, LN+; or (a5) ER+, ER/AGE low, LN^-;

(b) determining a profile comprising measurements of a plurality of genes or markers in a cell sample taken from said breast cancer patient, said plurality of genes markers comprising at least two of the genes or markers corresponding to the markers in (b1) Table 1 if said breast cancer patient is classified as ER^-, *BRCAI*; (b2) Table 2 if said breast cancer patient is classified as ER^- sporadic; (b3) Table 3 if said breast cancer patient is classified as ER+, ER/AGE high; (b4) Table 4 if said breast cancer patient is classified as ER+, ER/AGE low, LN+; or (b5) Table 5 if said breast cancer patient is classified as ER+, ER/AGE low, LN^-; and

(c) classifying said breast cancer patient as having a good prognosis or a poor prognosis based on said profile of said plurality of genes or markers,

wherein ER^+ designates a high ER level and ER^- designates a low ER level, wherein said ER/AGE is a metric of said ER level relative to the age of said patient, and wherein LN^+ designates a greater than 0 lymph nodes status in said patient and LN^- designates a 0 lymph nodes status in said patient.

43. The method of claim 42, wherein step (c) is carried out by a method comprising comparing said profile to a good prognosis template and/or a poor prognosis template, and wherein said patient is classified as having a good prognosis if said profile has a high similarity to a good prognosis template or has a low similarity to a poor prognosis template or as having a poor prognosis if said profile has a low similarity to a good prognosis template or has a high similarity to a poor prognosis template, said good prognosis template comprising measurements of said plurality of genes or markers representative of levels of said genes or markers in a plurality of good outcome patients and said poor prognosis template comprising measurements of said plurality of genes or markers representative of levels of said genes or markers in a plurality of poor outcome patients, wherein a good outcome patient is a breast cancer patient who has non-reoccurrence of metastases within a first period of time after initial diagnosis and a poor outcome patient is a patient who has reoccurrence of metastases within a second period of time after initial diagnosis.

44. The method of claim 43, further comprising determining said profile, said ER level, said LN status, and/or, said ER/AGE.

45. The method of claim 44, wherein said profile is an expression profile comprising measurements of a plurality of transcripts in a sample derived from said patient, wherein said

good prognosis template comprises measurements of said plurality of transcripts representative of expression levels of said transcripts in said plurality of good outcome patients, and wherein said poor prognosis template comprises measurements of said plurality of transcripts representative of expression levels of said transcripts in said plurality of poor outcome patients.

46. The method of claim 45, wherein said expression profile is a differential expression profile comprising differential measurements of said plurality of transcripts in said sample derived from said patient versus measurements of said plurality of transcripts in a control sample.

47. The method of claim 43, wherein said profile comprises measurements of a plurality of protein species in a sample derived from said patient, wherein said good prognosis template comprises measurements of said plurality of protein species representative of levels of said protein species in said plurality of good outcome patients, and wherein said poor prognosis template comprises measurements of said plurality of protein species representative of levels of said protein species in said plurality of poor outcome patients.

48. The method of claim 46, wherein measurement of each said transcript in said good prognosis template is an average of expression levels of said transcript in said plurality of good outcome patients.

49. The method of claim 48, wherein similarity of said expression profile to said good prognosis template is represented by a correlation coefficient between said expression profile and said good prognosis template, wherein said correlation coefficient greater than a correlation threshold indicates a high similarity and said correlation coefficient equal to or less than said correlation threshold indicates a low similarity.

50. The method of claim 48, wherein similarity of said expression profile to said good prognosis template is represented by a distance between said cellular constituent profile and said good prognosis template, wherein said distance less than a given value indicates a high similarity and said distance equal to or greater than said given value indicates a low similarity.

51. The method of claim 49, wherein said correlation threshold is 0.5.

52. The method of claim 51, wherein said ER level is determined by measuring an expression level of a gene encoding said estrogen receptor in said patient relative to expression level of said gene in said control sample, and wherein said ER level is classified as ER^+ if log10(ratio) of said expression level is greater than -0.65, and wherein said ER level is classified as ER^- if log10(ratio) of said expression level is equal to or less than -0.65.

53. The method of claim 52, wherein said gene encoding said estrogen receptor is the estrogen receptor α gene.

54. The method of claim 53, wherein said ER/AGE is classified as high if said ER level is greater than $c\cdot(AGE - d)$, and wherein said ER/AGE is classified as low if said ER level is equal to or less than $c\cdot(AGE - d)$, wherein c is a coefficient, AGE is the age of said patient, and d is an age threshold.

55. The method of claim 54, wherein said estrogen receptor level is measured by a polynucleotide probe that detects a transcript corresponding to the gene having accession number NM_000125, wherein said control sample is a pool of breast cancer cells of different patients, and wherein $c = 0.1$ and $d = 42.5$.

56. The method of claim 55, wherein said control sample is generated by pooling together cDNAs of said plurality of transcripts from a plurality of breast cancer patients.

57. The method of claim 55, wherein said control sample is generated by pooling together synthesized cDNAs of said plurality of transcripts and said transcript of said gene encoding said estrogen receptor.

58. The method of claim 42, wherein said individual is ER^-, *BRCA1*, and said plurality of genes comprises at least two of the genes for which markers are listed in Table 1.

59. The method of claim 42, wherein said individual is ER^-, *BRCA1*, and said plurality of genes comprises all of the genes for which markers are listed in Table 1.

60. The method of claim 42, wherein said individual is ER^-, sporadic, and said plurality of genes comprises at least two of the genes for which markers are listed in Table 2.

61. The method of claim 42, wherein said individual is ER^-, sporadic, and said plurality of genes comprises all of the genes for which markers are listed in Table 2.

62. The method of claim 42, wherein said individual is ER+, ER/AGE high, and said plurality of genes comprises at least two of the genes for which markers are listed in Table 3.

63. The method of claim 42, wherein said individual is ER+, ER/AGE high, and said plurality of genes comprises all of the genes for which markers are listed in Table 3.

64. The method of claim 42, wherein said individual is ER+, ER/AGE low, LN+, and said plurality of genes comprises at least two of the genes for which markers are listed in Table 4.

65. The method of claim 42, wherein said individual is ER+, ER/AGE low, LN+, and said plurality of genes comprises all of the genes for which markers are listed in Table 4.

66. The method of claim 42, wherein said individual is ER+, ER/AGE low, LN$^-$, and said plurality of genes comprises at least two of the genes for which markers are listed in Table 4.

67. The method of claim 42, wherein said individual is ER+, ER/AGE low, LN$^-$, and said plurality of genes comprises all of the genes for which markers are listed in Table 4.

68. The method of claim 42, wherein said profile further comprises one or more genes for which markers are not found in Tables 1-5, wherein said one or more genes are informative for prognosis.

69. A method for assigning an individual to one of a plurality of categories in a clinical trial, comprising assigning said individual to one category in a clinical trial if said individual has a good prognosis as determined by the method of any one of claims 7-33, and assigning said individual to a second category in said clinical trial if said individual has a poor prognosis as determined by the method of any one of claims 7-33.

70. The method of claim 69, wherein said individual is additionally assigned to a category in said clinical trial on the basis of the classification of said individual as determined in step (a).

71. The method of claim 69, wherein said individual is additionally assigned to a category in said clinical trial on the basis of one or more other clinical, phenotypic or genotypic characteristic of breast cancer.

72. The method of claim 69, further comprising determining in said cell sample the levels of expression of said one or more genes for which markers are not found in Tables 1-5, and determining from said expression levels of said one or more genes, whether said individual has a good prognosis or a poor prognosis.

73. A microarray comprising a plurality of polynucleotide probes each complementary and hybridizable to a sequence in a different gene listed in any one of Tables 1-5.

74. A microarray comprising a plurality of polynucleotide probes each complementary and hybridizable to a sequence in a different gene listed in Table 1.

75. The microarray of claim 74, wherein said plurality of polynucleotide probes comprises a probe complementary and hybridizable to each of the genes listed in Table 1.

76. A microarray comprising a plurality of polynucleotide probes each complementary and hybridizable to a sequence in a different gene listed in Table 2.

77. The microarray of claim 76, wherein said plurality of polynucleotide probes comprises a probe complementary and hybridizable to a sequence in each of the genes listed in Table 2.

78. A microarray comprising a plurality of polynucleotide probes each complementary and hybridizable to a sequence in a different gene listed in Table 3.

79. The microarray of claim 78, wherein said plurality of polynucleotide probes comprises a probe complementary and hybridizable to a sequence in each of the genes listed in Table 3.

80. A microarray comprising a plurality of polynucleotide probes each complementary and hybridizable to a sequence in a different gene listed in Table 4.

81. The microarray of claim 80, wherein said plurality of polynucleotide probes comprises a probe complementary and hybridizable to a sequence in each of the genes listed in Table 4.

82. A microarray comprising a plurality of polynucleotide probes each complementary and hybridizable to a sequence in a different gene listed in Table 5.

83. The microarray of claim 82, wherein said plurality of polynucleotide probes comprises a probe complementary and hybridizable to a sequence in each of the genes listed in Table 5.

84. The microarray of any of claims 73-83, wherein said plurality of polynucleotide probes constitutes at least 50% of the probes on said microarray.

85. The microarray of any of claims 73-83, wherein said plurality of polynucleotide probes constitutes at least 90% of the probes on said microarray.

86. The microarray of claim 73, wherein said plurality of polynucleotide probes comprises probes complementary and hybridizable to at least 75% of the genes listed in Table 1, Table 2, Table 3, Table 4, or Table 5, wherein said plurality of polynucleotide probes, in total, constitutes at least 50% of the probes on said microarray.

87. A kit comprising the microarray of any one of claims 73-83 in a sealed container.